

Distributed Flow Scheduling in an Unknown Environment

Yaoqing Yang ^{*†}, Keqin Liu [†], Qing Zhao [†]

^{*} Department of Electronic Engineering, Tsinghua University, Beijing, China

[†] Department of Electrical and Computer Engineering, UC Davis, California, USA

Email: yqyang1991@gmail.com

Abstract

Flow scheduling tends to be one of the oldest and most stubborn problems in networking. It becomes more crucial in the next generation network, due to fast changing link states and tremendous cost to explore the global structure. In such situation, distributed algorithms often dominate. In this paper, we design a distributed virtual game to solve the flow scheduling problem and then generalize it to situations of unknown environment, where online learning schemes are utilized. In the virtual game, we use incentives to stimulate selfish users to reach a Nash Equilibrium Point which is valid based on the analysis of the ‘Price of Anarchy’. In the unknown-environment generalization, our ultimate goal is the minimization of cost in the long run. In order to achieve balance between exploration of routing cost and exploitation based on limited information, we model this problem based on Multi-armed Bandit Scenario and combined newly proposed DSEE with the virtual game design. Armed with these powerful tools, we find a totally distributed algorithm to ensure the logarithmic growing of regret with time, which is optimum in classic Multi-armed Bandit Problem. Theoretical proof and simulation results both affirm this claim. To our knowledge, this is the first research to combine multi-armed bandit with distributed flow scheduling.

Keywords—*Flow Scheduling, Price of Anarchy, Multi-Armed Bandit, Logarithmic Regret*

I. INTRODUCTION

We consider a network sharing optimization problem. All of the users would like to optimize their own path selection without exchanging information with others. However, congestion on the same edge introduces increasing cost. We would like to figure out a distributed scheme for them to find a best solution.

We assume here that each user has a flow with unit capacity requirement but different source or destination. However, generalization to multi-commodity situation is not difficult if we split flows into units and carry out the algorithm for each unit flow. Cost on each edge is a random variable due to link state changes and environment variances. As mentioned above, conflictions increase costs, so we assume the expectation of one such variable grows when flows routed on it increase. In the front half of this paper, we assume these expectations are known and we focus on the virtual game designing to find the flow scheduling scheme.

In the second half, we generalize our problem into unknown environment. That is, we do not know the expectations of edge costs and we need moderate exploration. We use the newly proposed DSEE Sequence[17] to optimize the time for exploration. After exploration, samples of edge costs are stored in routers and the sample means are calculated to approximate the expectations. Exploration periods happen periodically in a predetermined manner so routers know when to explore. Between two neighboring exploration periods is an exploitation period. At the beginning of an exploitation period, we use the distributed Bellman-Ford algorithm[16] to calculate routing tables based on the sample means. In order to solve the confliction problem, we apply the virtual game here. During the rest time of the exploitation period, we route flows according to the routing tables. Obviously, exploration and Bellman Ford periods both introduce extra cost, or reward loss. The ultimate object

for us is to design a distributed algorithm to minimize long-run total cost for the whole network. In the whole paper, we assume that time is slotted and both explorations and exploitations need time.

A. *Background of Flow Scheduling*

Problems of flow scheduling in known scenario could still be very hard to solve. There are increasing literatures in this area with development of the widely-used MPLS network. Here, we base our work on background of flow scheduling instead of packet switching, wired or wireless, in order to make it more practical and useful nowadays.

The *minimum interference routing* [1]-[4] is a prospective direction in flow scheduling. Its purpose can be quite similar with ours. However, minimum interference routing algorithms, like MIRA[2] and WSP[4], consider more about load balancing to maintain the sustainability of future flow admitting, while we want to solve an optimization problem right now. Extensions of our work approve of adaptive scheduling of newly admitted flow but all routers should be informed beforehand that new flows have come in.

Literatures in the Routing Games are more relevant to our problem. Firstly, our modeling is very similar to the modeling of the *atomic routing* in [5]. Secondly, at the Bellman Ford period users perform a virtual game and take turns to select their own optimized routing path without considering congestion to others, which is the same with routing games. However, there is still fundamental difference between our virtual game and atomic routing. Firstly, we let distributed routers decide the best paths for the players, other than players select by themselves. This is more reasonable since in real life, routers decide paths for users. Secondly, our game is only virtual, which is used finally to solve an optimization problem. However, it is well known that games won't always converge to the optimum point. So we set the extra cost one user introduces to the whole network as the revenue he pays (see part II.B) to make this non-cooperative game a situation when selfish optimization equals social optimization. We prove the fast convergence to Nash Equilibrium Point in this routing game and use the constant bound of the 'Price of Anarchy' to measure its worth[9]. Moreover, modeling of [5] does not consider the generalization to unknown environment, so our work is more general.

B. *Stochastic online learning based on MAB Problem*

Second half of our paper focuses on the generalization to unknown model. The nature of routing problem with unknown edge cost calls for introduction of the Multi-armed Bandit (MAB) Problem. In the classic MAB, there are N independent arms and one single player. Each arm, when played, incurs a random cost with an unknown distribution. The player should decide the sequence to play each arm to obtain the minimum cost. We notice that the player should try to maintain the balance between exploration and exploitation, which respectively means to play a new arm and learn its cost distribution and to play the arm with minimum cost. A frequently used criterion to judge the performance of an adopted sequence is the so called *regret* or *cost of learning*, defined as the difference in total cost between the chosen sequence and the optimum sequence when cost distribution is known. The best regret, logarithmically growing with time, is obtained in [10] by Lai and Robbins. In [11][12], authors gave out index-type policies to achieve logarithmic regret.

Routing problems with unknown edge cost distributions can be modeled as a variation of the classic MAB problem if we view each path as an arm. However, performances of classic algorithms degrade severely here since paths with shared edge cannot be viewed as independent. In [13], Liu and Zhao explore the dependence of paths to obtain a logarithmic regret. In [14], Gai and Krishnamachari made modifications to UCB1 [12] and applied their algorithm LLC into shortest path problem. However, none of them gave out distributed method for path selection. In our work, we put this difficulty into the design of a distributed virtual game and solve it beforehand in known model. It's important to note that the concept Distributed Learning in [15] is different from our concept of 'distributed'. 'Distributed' in [15] means that each user does not exchange information with others and finds the best arm on his own. However, we further assume that our algorithm should be carried out

distributedly in each router by using the Bellman Ford Algorithm. Moreover, [12]-[15] did not consider network sharing, so our work is more general.

In our paper, we explore an algorithm doing online learning for multi-user situation in a distributed way. To our knowledge, no previous work considered such comprehensive situation. Based on our algorithm, the whole network can also achieve logarithmic regret with time. However, in order to judge the virtual game at the same time, we define regret slightly differently from the classic definition.

Definition 1: We define **Regret** as the number of time slots when the network is not in a Nash Equilibrium Point.

In Regret Analysis part, we analyze the equivalence between definition 1 and the classic one. We prove that our virtual game reaches a Nash Equilibrium Point in limited circles, and regret grows logarithmically with time. These claims ensure the effectiveness of the virtual game.

It is important to note that the Optimum Point is also a Nash Equilibrium Point in our game. However, Nash Equilibrium Point is not unique since strategy domain for each user is discrete (different paths). Commonly, only when we have continuous strategy domain, Nash Equilibrium Point is unique[5][6]. So analysis of the *Price of Anarchy* is necessary.

II. SYSTEM MODEL

A. Cost Modeling

Consider a graph $G = (V, E)$ and K source-destination pairs (s_k, t_k) , each with unit amount $f_k = 1$. For each edge $e \in E$, define flow on the edge

$$f_e = \sum_{e \in p_k} f_k \quad (1)$$

in which the p_k represents the path chosen by the k th flow. Since all flows have unit amount, the flow on each edge f_e will take discrete value from $\{1, 2, \dots, K\}$. Define

$$C(F) = \sum_{e \in E} c_e(f_e) \quad (2)$$

as the total cost in one time slot, in which the c_e represents the cost for edge e . At each time slot, for each edge e and a certain flow amount f_e , $c_e(f_e)$ is a random variable whose expectation value increases when f_e grows. For different time slots, $c_e(f_e)$ is an i.i.d. random process. F denotes the whole flow distribution on the network. In order to minimize the time average of $C(F)$, we try to obtain the best flow distribution F in a distributed way to minimize the expectation of $C(F)$. Henceforth we use a bar to represent the expectation. For example, $\bar{C}(F)$ denotes the expectation of $C(F)$. The unit amount is the granularity of all flows. Obviously, generalization to multi-commodity scenario is trivial if we split flows into flow units and treat each unit as an independent flow.

B. Incentive

In the virtual game design, users are assumed selfish since they could not exchange information. In order to stimulate users to cooperate, we set revenues as incentives for them. Assume at some time t , there are already K_t flows in the network and the whole flow distribution is currently F_t . Then the whole cost of the network equals $\bar{C}(F_t)$. For a certain k th flow, let $F_t(k)$ denote the flow distribution when f_k is withdrawn from F_t . Then we define

$$\bar{C}(F_t) - \bar{C}(F_t(k)) \quad (3)$$

as the revenue for the k th flow. We can easily see that when a user has the opportunity to change its routing path, he surely chooses the path that introduces the minimum extra cost to the whole network. Then the total cost decreases.

III. ALGORITHM IN KNOWN MODEL

A. Virtual Game Design

In this part, we assume that routers know all $\bar{c}_e(f_e)$ beforehand. Each user takes turns to hire routers to do Bellman Ford Algorithm. The price for each edge is set as the incentive described in II.B.

There will be $N * K$ time slots reserved for one circle. So time reserved for each user is N slots, and the Bellman Ford Algorithm surely converges in such long period. Also, total cost decreases each time when a user changes path, since the revenue for this user defined earlier is equal to the extra cost to the whole network introduced by him.

The complete algorithm is as follows:

- 1) Take out the k th flow from current flow distribution. If it is the first time for this flow to do path optimization and routers do not know yet the path to transmit this flow, they do not need to take it out.
- 2) Calculate price on each edge. The price is the extra cost if this edge is chosen:

$$P_e(F) = \bar{c}_e(f_e) - \bar{c}_e(f_e - f_k) \quad (4)$$

- 3) Start the Bellman Ford Algorithm and wait for N slots to ensure its convergence. The source node is s_k . Find out the path with minimum price to transmit flow to d_k .

- 4) Add up f_k on each edge chosen to transmit the k th flow.

- 5) Do the 1) again for the $k + 1$ th flow.

B. Nash Equilibrium Point

Theorem 1: If we do algorithm described in III.A, then after finite circles, the whole network reaches a Nash Equilibrium Point. Convergence time is bounded.

Proof: During one circle, one of two events below must occur:

- a). At least one user changes his path.
- b). No one changes his path.

If event ‘b’ happens, we know that no one could change his path unilaterally. Obviously the network has reached Nash Equilibrium Point.

However, if event ‘a’ happens, total cost decreases. This has been stated in II.B. Since there will be limited paths for one flow to take, number of flow distribution is limited, too. So ‘a’ won’t happen all the time.

We can further figure out the upper bound of convergence time to reach a Nash Equilibrium Point. In fact, we need $\lceil \frac{S_M}{S_m} \rceil$ times of Bellman Ford circles. The S_M denotes the maximum difference between cost of two different flow distribution, and S_m denotes the minimum. This is true because during each Bellman Ford circle, the cost of the whole network will at least decrease by S_m if ‘b’ does not happen. \square

IV. PRICE OF ANARCHY

In this part we give out the analysis of the ‘Price of Anarchy’. This notion was originally defined in [8] to measure the selfish performance of a simple game of N players that compete for M parallel links. In [9], the authors analyzed the price of anarchy of an atomic routing game to polynomial edge cost with nonnegative coefficient. They gave out results of $d^{O(d)}$ in which d represents the highest order of the polynomial edge cost function. This result is considered by [5] to be a significant generalization of previous work.

In our paper, we still need analysis of the ‘Price of Anarchy’ since our ultimate goal is to solve an optimization problem. So far, we give out algorithm to make different users optimize their own price—the incentive—to reach a Nash Equilibrium Point. So we need to figure out the difference between a Nash Equilibrium Point and the

optimum point.

Definition 2: We define the **Price of Anarchy** as

$$\bar{C}(F_N)/\bar{C}(F^*) \quad (5)$$

The F_N represents flow distribution of one Nash equilibrium point. And F^* represents flow distribution of the optimum point, in which (2) is optimized.

We give out existence of constant price of anarchy for general polynomial edge cost. Then we give out concrete value for polynomials with nonnegative coefficients. Here we need the functions to be convex but this is trivial when congestion is concerned. The assumption of polynomial edge cost is common in previous work of Routing Games[5][8][9]. Our modeling is different from routing game. In Routing Games, the ‘total price’ in (6) is equal to the expectation of total cost function defined in (2), while they are different in our virtual game. However, polynomial functions are quite enough to model congestions in our problem, so we still use this assumption.

A. General Polynomial Function: Existence

In this part we prove the existence of constant upper bound of the price of anarchy for polynomial edge cost function. In another word, this constant is independent of network size and topology. In the proof we use the following definition.

Definition 3: We define **Total Price** for distribution F as

$$P(F) = \sum_{e \in E} [\bar{c}_e(f_e) - \bar{c}_e(f_e - f_u)] \cdot f_e \quad (6)$$

f_u just means the unit flow amount.

We simply replace f_u with 1 in following parts, since we claim in section II.A that all flow has the same unit amount. What is important is the reason we define (6) as the ‘total price’. In fact, from (3)(4) we know that the incentive pricing scheme asks for the k th user a price of

$$P_k(F) = \sum_{e \in p_k} [\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] \quad (7)$$

We add up (7) for all users and we get

$$P(F) = \sum_{k=1}^K \sum_{e \in p_k} [\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] \quad (8)$$

Simply change the order of summation and we get (6).

Theorem 2: If the expectation of edge cost function $\bar{c}_e(f_e)$ is convex and grows polynomially with f_e , there exists a constant bound for the ‘Price of Anarchy’ independent of network size and flow amount.

We assume that edge cost functions are polynomials of maximum degree d . Here d is different from the degree of barycentric spanner in proof of **Theorem 4**.

$$\bar{c}_e(f_e) = a_e f_e^d + \sum_{i=1}^d a_e^{(i)} f_e^{d-i} \quad (9)$$

First we give out Lemma 1. This is the relationship between total cost and total price.

Lemma 1: For a given network $G=(V,E)$, there exist two constant numbers A_l, A_r . For any flow distribution F , we have

$$A_l \leq \frac{P(F)}{C(F)} \leq A_r \quad (10)$$

These two numbers are independent of the network size.

The nature of Lemma 1 is very simple. For a polynomial $E(c_e)$, the numerator and the denominator of (10) is of the same order of flow amount f_e . So the fraction is certainly limited. We put detailed proof in Appendix A. Similarly, we could arrive at the following formula.

For a given $G=(V,E)$, there exists a constant number A_u . For any flow distribution and any edge e , it satisfies

$$\frac{\bar{c}_e(f_e + 1) - \bar{c}_e(f_e)}{\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)} \leq A_u \quad (11)$$

Then we give out **Lemma 2**. This is the ‘Variational Inequality Characterization’[5], which describes the basic feature of a Nash Equilibrium Point. Proof of **Lemma 2** is also put in the appendix.

Lemma 2: For a given network $G=(V,E)$ and a Nash Equilibrium point F of K users, for any flow distribution F' , we have

$$\sum_{e \in E} [\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] \cdot f_e \leq A_u \sum_{e \in E} [\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] \cdot f_e' \quad (12)$$

Based on these two Lemmas, we can complete the proof of **Theorem 2**. The proof is still very simple in nature. We have proven that the total cost(2) and the total price(6) grows with flow amount in the same order (**Lemma 1**). Then we find the constant upper bound of $\frac{P(F_N)}{P(F^*)}$ (**Lemma 2**). These two steps complete the proof. The detailed proof is put in Appendix C.

B. Polynomial Function with Nonnegative Coefficients: Concrete Value

For polynomial edge cost with nonnegative coefficients, we give out concrete value of the upper bound. Although we could derive a proof based on the same procedure of part IV.A, we can take advantage of the nonnegative coefficients to get a relatively simple proof in the Appendix. First we give out some definitions. If (8) holds and coefficients are all nonnegative, we have for each edge e

$$\bar{c}_e(f_e + 1) - \bar{c}_e(f_e) = a_e[(f_e + 1)^d - f_e^d] + \sum_{i=1}^d a_e^{(i)}[(f_e + 1)^{d-i} - f_e^{d-i}] \quad (13)$$

Obviously, all terms in (13) have nonnegative coefficients. We assume

$$\bar{c}_e(f_e + 1) - \bar{c}_e(f_e) = \sum_{i=0}^d \tilde{a}_e^{(i)} f_e^{d-i-1} \quad (14)$$

in which $\tilde{a}_e^{(i)} > 0$ and $\tilde{a}_e^{(0)} = a_e$. Moreover,

$$\sum_{i=0}^d \tilde{a}_e^{(i)} = \bar{c}_e(1) - \bar{c}_e(0) \quad (15)$$

We assume

$$s_e = \min_{a_e^{(i)} > 0} (a_e^{(i)}) \quad (16)$$

$$L = \max_{e \in E} [\bar{c}_e(1) - \bar{c}_e(0)] \quad (17)$$

Theorem 3 For a given network $G=(V,E)$, if all edge cost functions satisfy (9) and coefficients are nonnegative, we could give out the concrete value of the constant upper bound of the Price of Anarchy. The constant is $[(d+1)L \max_{e \in E} \frac{1}{s_e}]^d = d^{O(d)}$.

V. ALGORITHM IN UNKNOWN MODEL

From this section, we give out generalization to unknown model. In another word, we further assume that the cost distribution of each edge is unknown at the beginning. In order to get enough information about the network, we adopt the newly proposed DSEE Sequence algorithm in [17] and cut time into interleaving exploration and exploitation periods. A router sends exploration flows to get samples of the cost and store them in memory. Based on these samples, a router calculates sample mean and view it as the expectation of edge cost when doing Bellman-Ford Algorithm. Between the exploration periods are the exploitation periods, at the beginning of which the virtual game is applied. During the rest time of exploitation, users share the network based on routing tables. In order not to route flows on edges with high price, each user consents to do enough explorations. However, exploration periods and Bellman Ford periods cannot be too long since they introduce extra cost to the network.

A. Exploration

One exploitation period lasts for $N = |V|$ time slots. In one exploration period, only one source node s_k starts exploration. K source nodes take turns to do exploration in different exploration periods. At the beginning of the first exploration period, s_1 sends out a short flow of a random amount k_1 to a random edge e_r related to it to explore the value $c_{e_r}(k_1)$. Then the other node of edge e_r receives this flow and forward it in the next time slot. This whole exploration period terminate in $N = |V|$ time slots. In the next exploration period, the source node s_2 starts exploration instead of s_1 . The constant number $N = |V|$ is large enough to ensure a minimum probability $r = \min_{e \in E} (r_e) > 0$, in which the r_e is the probability of the edge e being estimated.

B. Exploitation

At the beginning of this period, there will be $N * K$ time slots reserved for a Bellman Ford period. During one period, we do one circle of the virtual game described in III.A.

However, we should replace (4) with

$$P_e(F) = \hat{c}_e(f_e) - \hat{c}_e(f_e - f_k) \quad (18)$$

in which $\hat{c}_e(f_e)$ denotes the sample means stored in routers' memory.

C. DSEE

Time is divided into interleaving sequence of Exploration and Exploitation. At the beginning of each exploitation period, there is $N * K$ time slots arranged for Bellman Ford period to do virtual game. One Bellman Ford period terminates only when the total time $N * K$ is reached. Similarly, one Exploration period ends after N time slots. However, the exploitation period ends when the time slot t satisfies

$$card(t) < G \log(t) \quad (19)$$

in which the $card(t)$ represents number of time slots used to do exploration up to time t . Certainly, the whole DSEE Sequence is determined beforehand once the parameter G has been chosen.

VI. REGRET ANALYSIS

We define regret as the number of time slots when all the flows are not routed in Nash Equilibrium Point (see the end of the Introduction part). In section III.B, we have proved the inevitability for K users to reach the Nash Equilibrium Point in limited circles of virtual game. In this part, we analyze the equivalence of definition 1 with classic one. Then we prove regret grows logarithmically with time.

A. Equivalence between Definition 1 and classic definition

Classic definition of regret is the difference in total cost between the chosen strategy sequence and the optimum strategy sequence when cost distribution is known.

In our algorithm, there exist two conditions that regret increases. The first one is exploration or Bellman Ford. During these periods, no flows are transmitted. However, if we define an extra constant cost for each of such slot to get a classic definition, we can see that this two regrets grow with time in the same order. The second one is when flows are not routed in a Nash Equilibrium Point in an exploitation period. But in one such slot, extra cost cannot be larger than S_M . Therefore, even if we define a classic regret, it still grows with same order of time.

The only difference is the distance from one Nash Equilibrium Point to the Optimum Point. However, finding the Optimum Point for different flows tends to be NP hard and it cannot be done in a distributed way. So we choose to define regret based on a sub-optimal Nash Equilibrium Point which cannot be further improved in a distributed manner. Previous parts have shown the constant ‘Price of Anarchy’ bound, which convince of the feasibility of our definition.

B. Regret Order

Theorem 4: If the chosen G in (19) satisfies

$$G \geq \max(3/r, \frac{8d^2|E|\sigma^2}{rc^2}) \quad (20)$$

then $\text{regret}(T)$ increases with the form $O(\log(T))$.

Here we give out some definitions in Theorem 4.

Definition 4: Let S be a d -dimensional vector space. A set $B = \{x_1, x_2, \dots, x_d\} \subset S$ is called a barycentric spanner for S if every x in S can be written as linear combination of elements of B with coefficients in $[-1, 1]$.

It is shown in [15] that if S is a compact set, then it has a barycentric spanner. We know that the set of different paths for a certain source-destination pair (s_k, d_k) is a compact vector space, thus it has a barycentric spanner with dimension d_k . We assume $d = \max_{k=1 \sim K} d_k$. σ^2 is the largest variance of all the edge cost under different flow distributions. r is the minimum of the probability that a certain edge is chosen during explorations. c_k is the minimum price difference between two paths for the k th user under all different flow distributions. Since number of flow distributions is limited, c_k surely exists. Then we can define $c = \min_{k=1 \sim K} c_k$. These parameters are all related to the network topology and can be obtained beforehand. However, while choosing a G based on (20) is doable, usually we can choose a smaller G . Here we only concern about the existence of a sufficient condition.

Proof of **Theorem 4** still can be found in the Appendix. Instead we give out the basic idea of the proof. If G is chosen big enough, sufficient times will be used for exploration so that we have relatively accurate sample means for the cost of each edge under different flow amount. Based on Bernstein’s inequality, we can bound the variance of sample means of path cost. When this variance is small enough, we can bound the probability that

we make mistakes in the virtual game circle. Mistake-free virtual game results in Nash Equilibrium. Although proof of **Theorem 4** seems lengthy, it relies on this simple idea.

VII. SIMULATIONS

A. Price of Anarchy Simulation

In this part we give out simulation result for the ‘Price of Anarchy’. Figure 1 shows the probability density function of the ‘Price of Anarchy’ for different cost function orders. Large density near price 1 proves the efficiency of our algorithm. Also, the relationship between the ‘Price of Anarchy’ and cost function order can be observed: distribution with a higher order has a longer tail.

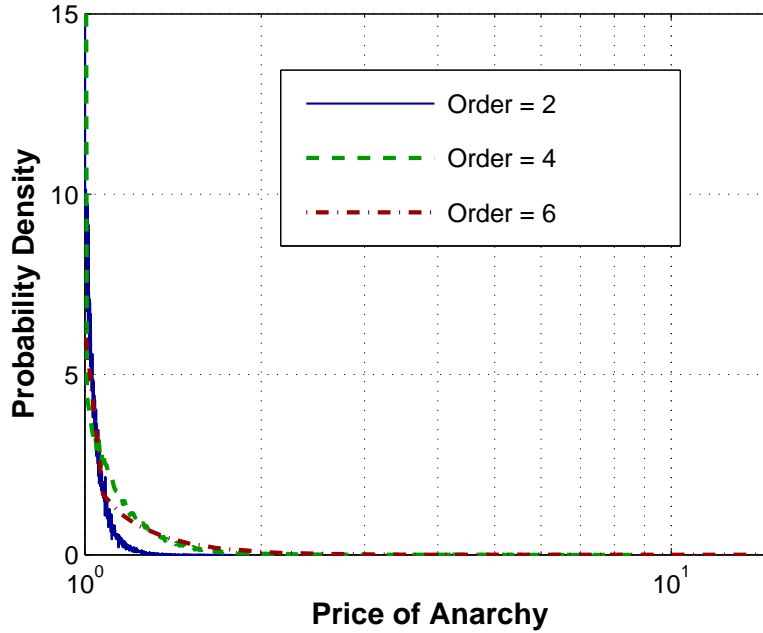


Fig. 1: ‘Price of Anarchy’ distribution

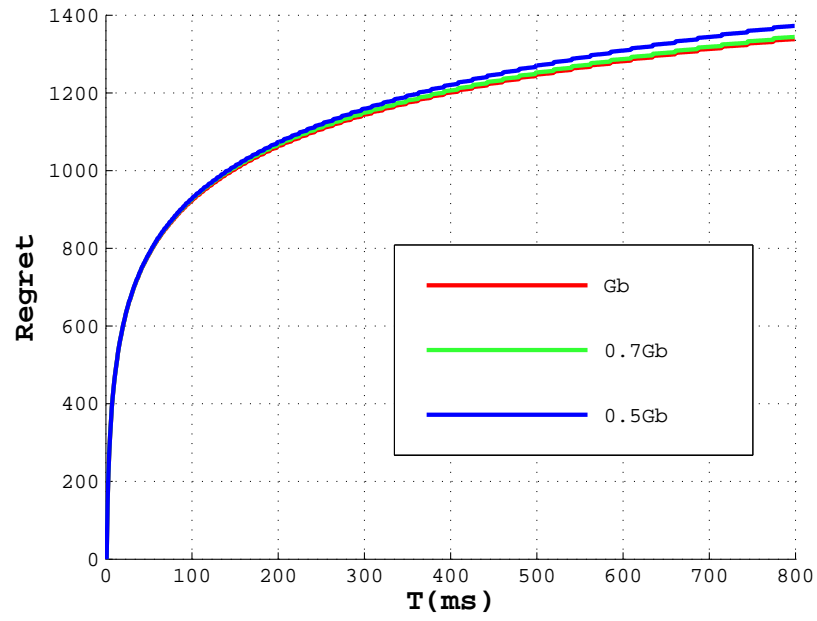
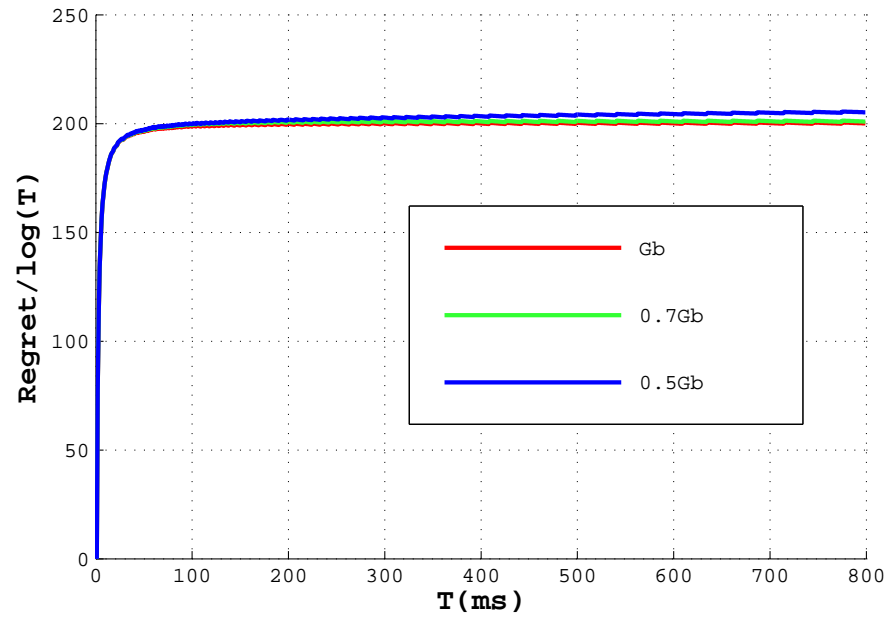
B. Regret Simulation

In this part we give out the simulation results for regret order. Figure 2 shows the growing behavior of regret with time under different G selections. We choose the G_b as the basic G based on the condition shown in **Theorem 4**. Actually, this condition is just an sufficient condition that leads to logarithmic growing of regret. In real simulation, we have chosen a basic G smaller than in **Theorem 4** but can still help the logarithmic growth hold.

The second figure is the regret divided by $\log(T)$. It could help us see more clearly how the regret converges to a logarithmic order. Moreover, we see from simulation that if G is not large enough, the regret grows with an order larger than $\log(T)$. So in real-life applications, we should make sure that G is large enough.

VIII. CONCLUSIONS

In this paper, we considered the flow scheduling problem both under known and unknown model. For the known model, we proposed a virtual non-cooperative game with incentive pricing to solve cost optimization problem for users who do not exchange information with each other. To analyze this virtual game, we proved

Fig. 2: regret(T)Fig. 3: regret(T) divided by $\log(T)$

the fast convergence of the game into a Nash Equilibrium Point which had a bounded price of anarchy. The constant bound was proved to be independent of network size and flow amount. Then we extended this algorithm to situations when cost distributions were unknown beforehand. We modeled this problem under multi-armed bandit model and combined the virtual game with the newly proposed DSEE Sequence which could achieve best regret for all light-tail cost distributions. Sure enough, regret of our algorithm was proved to be growing logarithmically with time if the DSEE parameters were chosen properly, which is best in the classic online learning scenario. Also, simulation results of the ‘Price of Anarchy’ and the regret growing behavior were given out to test the essential correctness of all our claims.

APPENDIX A PROOF OF LEMMA 1

Based on (9) we have

$$[\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] \cdot f_e = a_e f_e^d + \sum_{i=1}^d b_e^{(i)} f_e^{d-i} \quad (21)$$

Here $a_e^{(i)}$ and $b_e^{(i)}$ are coefficients. We do not require them to be nonnegative here, but in **Theorem 3**, we require $a_e^{(i)}$ to be nonnegative. Divide (21) with f_e^d and we get

$$\frac{\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)}{f_e^{d-1}} = a_e + \sum_{i=1}^d b_e^{(i)} f_e^{-i} \quad (22)$$

For any $\epsilon > 0$, there exists a $f_{e,\epsilon}$. For any $f_e > f_{e,\epsilon}$,

$$\left| \frac{\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)}{f_e^{d-1}} - a_e \right| < \epsilon \quad (23)$$

So we have, for any $f_e > f_{e,\epsilon}$,

$$a_e - \epsilon < \frac{\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)}{f_e^{d-1}} < a_e + \epsilon \quad (24)$$

Since $f_{e,\epsilon}$ is limited, there exists a closed section I_e . For any $f_e \leq f_{e,\epsilon}$,

$$\frac{\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)}{f_e^{d-1}} \in I_e \quad (25)$$

Since $\frac{\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)}{f_e^{d-1}} > 0$, $0 \notin I_e$. We choose $\epsilon < \frac{a_e}{2}$, and denote $J_e = I_e \cup [\frac{a_e}{2}, \frac{3a_e}{2}]$ and we have for any f_e ,

$$\frac{\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)}{f_e^{d-1}} \in J_e \quad (26)$$

Similarly, we divide (9) with f_e^d and finally get

$$\frac{\bar{c}_e(f_e)}{f_e^d} \in J'_e \quad (27)$$

Here J_e and J'_e are both closed sections excluding zero. Then for any flow distribution F , we have

$$\begin{aligned} \frac{P(F)}{\bar{C}(F)} &= \frac{\sum_{e \in E} [\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] \cdot f_e}{\sum_{e \in E} \bar{c}_e(f_e)} \\ &= \frac{\sum_{e \in E} [\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] / f_e^{d-1}}{\sum_{e \in E} \bar{c}_e(f_e) / f_e^d} \end{aligned} \quad (28)$$

From (26)(27) we know there exist two numbers A_l, A_r , for any flow distribution, (10) holds. \square

APPENDIX B
PROOF OF **LEMMA 2**

For a certain $k \in \{1, 2, \dots, K\}$, the Nash Equilibrium Point F satisfies

$$\sum_{e \in p_k^N} [\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] \leq \sum_{p_k \in \Gamma_k} \sum_{e \in p_k} [\bar{c}_e(f_e + 1) - \bar{c}_e(f_e)] \cdot f_{p_k} \quad (29)$$

Here Γ_k represents the set of all paths available to the k th user. And $f_{p_k} = 1$ only when the path $p_k \in \Gamma_k$ is chosen by the k th user. Otherwise it equals zero. Obviously, (29) can be derived directly from the definition of Nash Equilibrium Point. For different path selection schemes, f_{p_k} varies. However, (29) always holds. For a certain flow distribution F , we add up (29) for all K users and get

$$P(F_N) \leq \sum_{e \in E} [\bar{c}_e(f_e + 1) - \bar{c}_e(f_e)] \cdot f_e' \quad (30)$$

Since we have (11) already, we can get (12). \square

APPENDIX C
PROOF OF **THEOREM 2**

Let F represents a random Nash Equilibrium point and F^* denotes the optimum point. For a certain edge e , from (26), we have

$$\frac{[\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)]/f_e^{d-1}}{[\bar{c}_e(f_e^*) - \bar{c}_e(f_e^* - 1)]/(f_e^*)^{d-1}} \leq \frac{J_e^{(L)}}{J_e^{(R)}} \quad (31)$$

The $J_e^{(L)}$ and $J_e^{(R)}$ are left and right border of J_e . And $*$ represents the optimum point. From this inequality we can derive directly and get

$$\begin{aligned} [\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] \cdot f_e^* &\leq \left(\frac{J_e^{(L)}}{J_e^{(R)}}\right)^{\frac{1}{d}} \{[\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] \cdot f_e\}^{\frac{d-1}{d}} \\ &\quad \cdot \{[\bar{c}_e(f_e^*) - \bar{c}_e(f_e^* - 1)] \cdot f_e^*\}^{\frac{1}{d}} \end{aligned} \quad (32)$$

Based on the Hölder inequality, we get

$$\begin{aligned} \sum_{e \in E} [\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] \cdot f_e^* &\leq \left(\frac{J_e^{(L)}}{J_e^{(R)}}\right)^{\frac{1}{d}} \sum_{e \in E} \{[\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] \cdot f_e\}^{\frac{d-1}{d}} \\ &\quad \cdot \{[\bar{c}_e(f_e^*) - \bar{c}_e(f_e^* - 1)] \cdot f_e^*\}^{\frac{1}{d}} \\ &\leq \left(\frac{J_e^{(L)}}{J_e^{(R)}}\right)^{\frac{1}{d}} \left\{ \sum_{e \in E} [\bar{c}_e(f_e) - \bar{c}_e(f_e - 1)] \cdot f_e \right\}^{\frac{d-1}{d}} \\ &\quad \cdot \left\{ \sum_{e \in E} [\bar{c}_e(f_e^*) - \bar{c}_e(f_e^* - 1)] \cdot f_e^* \right\}^{\frac{1}{d}} \\ &= \left(\frac{J_e^{(L)}}{J_e^{(R)}}\right)^{\frac{1}{d}} \cdot [P(F)]^{\frac{d-1}{d}} \cdot [P(F^*)]^{\frac{1}{d}} \end{aligned} \quad (33)$$

Since (12) holds for every flow distribution F' , we could let $F' = F^*$ so

$$P(F) \leq A_u \cdot \left(\frac{J_e^{(L)}}{J_e^{(R)}}\right)^{\frac{1}{d}} \cdot [P(F)]^{\frac{d-1}{d}} \cdot [P(F^*)]^{\frac{1}{d}} \quad (34)$$

It means

$$P(F)/P(F^*) \leq (A_u)^d \cdot \frac{J_e^{(L)}}{J_e^{(R)}} \quad (35)$$

And we have Lemma 1, so we finally get

$$\begin{aligned}\bar{C}(F)/\bar{C}(F^*) &= \frac{\bar{C}(F)}{P(F)} \cdot \frac{P(F)}{P(F^*)} \cdot \frac{P(F^*)}{\bar{C}(F^*)} \\ &\leq (A_u)^d \cdot \frac{A_r J_e^{(L)}}{A_l J_e^{(R)}}\end{aligned}\quad (36)$$

From previous Lemmas, we know absolutely that constants on the right side of this inequality are independent from network topology and flow distribution. Since F_N is a random flow distribution, we have proved **Theorem 2**. \square

APPENDIX D PROOF OF **THEOREM 3**

Conditions in this theorem also ensure the functions are convex. So we have for any flow distribution

$$\bar{C}(F) \leq P(F) \quad (37)$$

From (15) we have, for any i and any $e \in E$

$$\tilde{a}_e^{(i)} < L \quad (38)$$

Based on *Hölder* inequality, we have

$$\begin{aligned}\sum_{e \in E} [\bar{c}_e(f_e + 1) - \bar{c}_e(f_e)] \cdot f_e^* &= \sum_{i=0}^d \sum_{e \in E} \tilde{a}_e^{(i)} f_e^{d-i-1} f_e^* \\ &\leq \sum_{i=0}^d \left\{ \sum_{e \in E} \tilde{a}_e^{(i)} (f_e^{d-i-1})^{\frac{d-i}{d-i-1}} \right\}^{\frac{d-i-1}{d-i}} \\ &\quad \cdot \left\{ \sum_{e \in E} \tilde{a}_e^{(i)} (f_e^*)^{d-i} \right\}^{\frac{1}{d-i}} \\ &\leq L \sum_{i=0}^d \left\{ \sum_{e \in E} \frac{1}{s_e} \bar{c}_e(f_e) \right\}^{\frac{d-i-1}{d-i}} \left\{ \sum_{e \in E} \frac{1}{s_e} \bar{c}_e(f_e^*) \right\}^{\frac{1}{d-i}} \\ &\leq L \max_{e \in E} \frac{1}{s_e} \cdot \sum_{i=0}^d \left\{ \bar{C}(F) \right\}^{\frac{d-i-1}{d-i}} \left\{ \bar{C}(F^*) \right\}^{\frac{1}{d-i}}\end{aligned}\quad (39)$$

Since $\bar{C}(F^*) \leq \bar{C}(F)$, we have

$$\sum_{e \in E} [\bar{c}_e(f_e + 1) - \bar{c}_e(f_e)] \cdot f_e^* \leq (d+1) L \max_{e \in E} \frac{1}{s_e} \cdot \left\{ \bar{C}(F) \right\}^{\frac{d-1}{d}} \left\{ \bar{C}(F^*) \right\}^{\frac{1}{d}} \quad (40)$$

For one random Nash equilibrium F and the optimum point F^* , from (30)(37) we have

$$\bar{C}(F) \leq P(F) \leq \sum_{e \in E} [\bar{c}_e(f_e + 1) - \bar{c}_e(f_e)] \cdot f_e^* \quad (41)$$

Combining (16)(17)(40)(41) we have

$$\bar{C}(F)/\bar{C}(F^*) \leq [(d+1) L \max_{e \in E} \frac{1}{s_e}]^d = d^{O(d)} \quad (42)$$

And this constant is independent of network topology and flow distribution. \square

APPENDIX E
PROOF OF **THEOREM 4**

Since the number of time slots used in exploration and Bellman Ford increases strictly with $O(\log t)$, we could only focus on the number of slots that all flows are not operating at the Nash equilibrium point. Define the A_t the event that all the flows are not operating at the Nash equilibrium point at time t . We give out the upper bound of $P(A_t)$.

Define B_t^k as the event that last Bellman Ford just before time slot t for the k th flow goes wrong since poor estimation of the path cost. Then

$$P(B_t^k) = P\{\hat{X}^*(t) \geq \min_{p \in P} \hat{X}_p(t)\} \quad (43)$$

The P denotes the set of paths that the k th flow can choose from. The $\hat{X}_p(t)$ is the incentive price for choosing path p . This price is calculated by adding up all the extra edge cost introduced by the k th flow. That is

$$\hat{X}_p(t) = \sum_{e \in E} \hat{c}_e(f_e) - \hat{c}_e(f_e - f_k) \quad (44)$$

The p^* represents the real best path for k th flow to choose if price expectation for each edge is known exactly. And the $\hat{X}^*(t)$ is the estimated price for choosing this path.

Let $n_e(k, t)$ be the number of times $e \in E$ is observed when the k units of flow are put on it up to time t during the exploration slots. Let $r_e(k)$ represents the probability that e with flow k on it is chosen to be observed at a random time slot. Since k can only take values from $\{1, 2, \dots, K\}$ and the number of edges is limited, we can ensure the existence of $r = \min_{e \in E} r_e$.

Obviously,

$$E(n_e(k, t)) = Gr_e(k) \log t \quad (45)$$

$$Var(n_e(k, t)) < Gr_e(k) \log t \quad (46)$$

so, based on Bernstein's inequality

$$\begin{aligned} P\{n_e(k, t) < \frac{1}{2} Gr \log t\} &\leq P\{n_e(k, t) < \frac{1}{2} Gr_e(k) \log t\} \\ &< \exp\left(-\frac{1}{2} \frac{E^2(n_e(k, t))}{\frac{1}{2} E(n_e(k, t)) + Var(n_e(k, t))}\right) \\ &= t^{-\frac{1}{3} Gr_e(k)} \leq t^{-1} \end{aligned} \quad (47)$$

Let $M = \frac{1}{2} Gr \log t$ and we can easily get

$$P\{\exists e \in E, k \in \{1, 2, \dots, K\}, s.t. n_e(k, t) < M\} < \sum_{e \in E, 1 \leq k \leq K} P\{n_e(k, t) < M\} < K|E|t^{-1} \quad (48)$$

We choose a barycentric spanner in the network and assume it has d_k elements $\{p_1, p_2, \dots, p_{d_k}\}$, then

$$\{\hat{X}^*(t) \geq \min_{p \in P} \hat{X}_p(t)\} \subseteq \{\hat{X}^*(t) - X^*(t) > \frac{c}{2}\} \cup_{l=1}^{d_k} \{\hat{X}_l(t) - X_l(t) < -\frac{c}{2d_k}\} \quad (49)$$

in which

$$X_l(t) = \sum_{e \in p_l} [\bar{c}_e(f_e + f_k) - \bar{c}_e(f_e)] \quad (50)$$

and $X^*(t)$ represents the real minimum expectation price of the path for k th flow.

Specifically for each p_l we have

$$\hat{X}_l(t) - X_l(t) = \sum_{e \in p_l} [\hat{c}_e(f_e + f_k) - \hat{c}_e(f_e)] - \sum_{e \in p_l} [\bar{c}_e(f_e + f_k) - \bar{c}_e(f_e)] \quad (51)$$

When enough times are used to estimate each edge, the value above will have a high probability to be small. Let L_l denote the number of edges in p_l . Then we have

$$\begin{aligned} & P\{\hat{X}_l(t) - X_l(t) < -\frac{c}{2d_k} | \forall e \in E, k \in \{1, 2, \dots, K\}, n_e(k, t) \geq M\} \\ & < P\{|\hat{X}_l(t) - X_l(t)| > \frac{c}{2d_k} | \forall e \in E, k \in \{1, 2, \dots, K\}, n_e(k, t) \geq M\} \\ & < \sum_{e \in p_l} P\{|\hat{c}_e(f_e) - \bar{c}_e(f_e)| > \frac{c}{2d_k L_l} \\ & \quad | \forall e \in E, k \in \{1, 2, \dots, K\}, n_e(k, t) \geq M\} \\ & + \sum_{e \in p_l} P\{|\hat{c}_e(f_e + f_k) - \bar{c}_e(f_e + f_k)| > \frac{c}{2d_k L_l} \\ & \quad | \forall e \in E, k \in \{1, 2, \dots, K\}, n_e(k, t) \geq M\} \\ & \leq 2L_l * 2exp(-\frac{1}{2} \frac{(\frac{c}{2d_k L_l})^2}{\frac{\sigma^2}{Grlogt}}) \\ & \leq 4|E|exp(-\frac{1}{2} \frac{(\frac{c}{2d|E|})^2}{\frac{\sigma^2}{Grlogt}}) \\ & \leq 4|E|t^{-1} \end{aligned} \quad (52)$$

Similar upper bound of $\hat{X}^*(t)$ can also be obtained. After that we get

$$P(B_t^k) < 4(|E| + |E|^2)t^{-1} + t^{-1} < 5|E|^2t^{-1} \quad (53)$$

Each event B_t^k leads to the event $A_{\tilde{t}}$ for some $\tilde{t} > t$. If we would like to make the whole K flows reach the Nash Equilibrium point, we should ensure that B does not happen for a period long enough before time t . In fact, if B does not happen, we will need $\lceil \frac{S_M}{S_m} \rceil$ circles of Bellman Ford period to do virtual game. This result is based on Theorem 1. This is because if B does not happen, it tends to be the same situation that routers know exactly the cost distribution of each edge.

The nature of DSEE Sequence makes the start point of each exploration period in an exponential sequence. We present this fact in a heuristic way. For the start time t_1 of a exploration period, we have

$$card(t_1) = Glogt_1 \quad (54)$$

and for the start point t_2 of the next exploration period we have

$$card(t_2) = Glogt_2 \quad (55)$$

Since $card(t_1) + NK = card(t_2)$, we have

$$\frac{t_2}{t_1} = exp(\frac{NK}{G}) \quad (56)$$

Let $\{t_1, t_2, \dots, t_{\lceil \frac{S_M}{S_m} \rceil}\}$ denote the starting points of last $\lceil \frac{S_M}{S_m} \rceil$ circles of Bellman Ford period before time t . And let $t_{\lceil \frac{S_M}{S_m} \rceil + 1}$ denote the starting point of the following period after time t . We see obviously that

$$\frac{t_{\lceil \frac{S_M}{S_m} \rceil + 1}}{t_1} = \exp\left(\frac{NK \lceil \frac{S_M}{S_m} \rceil + 1}{G}\right) \quad (57)$$

For any Bellman Ford time slot t^* between t_1 and $t_{\lceil \frac{S_M}{S_m} \rceil + 1}$, it satisfies that

$$\frac{t}{t^*} < \frac{t_{\lceil \frac{S_M}{S_m} \rceil + 1}}{t_1} = \exp\left(\frac{NK \lceil \frac{S_M}{S_m} \rceil + 1}{G}\right) \quad (58)$$

During these circles of Bellman Ford period, if B does not happen, the A_t does not happen either. So we have

$$\begin{aligned} P(A_t) &< \sum_{t^*, k=1,2,\dots,K} P(B_{t^*}^k) < \sum_{t^*, k=1,2,\dots,K} 5|E|^2(t^*)^{-1} \\ &< 10K|E|^2 \lceil \frac{S_M}{S_m} \rceil \exp\left(\frac{NK \lceil \frac{S_M}{S_m} \rceil + 1}{G}\right) t^{-1} \end{aligned} \quad (59)$$

In another word, the total regret to time horizon T can be written as

$$\sum_{t=1}^T P(A_t) = \sum_{t=1}^T O(t^{-1}) \quad (60)$$

and it is $O(\log T) \square$

REFERENCES

- [1] K. Kar, M. Kodialam, T. V. Lakshman, "Minimum Interference Routing of Bandwidth Guaranteed Tunnels with MPLS Traffic Engineering Applications", *IEEE JSAC*, vol. 18, no. 12, pp. 2566-2579, December 2000.
- [2] Mudi Kodialam T. V. Laksban, "Minimum Interference Routing with Applications to MPLS TraMic Engineering", *Proc.IEEE INFOCOM*, vol. 2, pp. 884-893, 2000.
- [3] B. Fortz, M. Thorup, "Internet Traffic Engineering by Optimizing OSPF Weights", *Proc.IEEE INFOCOM*, vol. 2, pp. 519-528, 2000.
- [4] R. Guerin, A. Orda, D. Williams, "QoS routing mechanisms and OSPF extensions", *Proc.IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 3, pp. 1903-1908, 1997.
- [5] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, "Algorithmic Game Theory", *Cambridge University Press*, 2007.
- [6] J. B. Rosen, "Existence and Uniqueness of Equilibrium Points for Concave N-Person Games", *Econometrica*, vol. 33, no. 3, pp. 520-534, Jul. 1965.
- [7] Baruch Awerbuch, Robert Kleinberg, "Online Linear Optimization and Adaptive Routing", *Journal of Computer and System Sciences*, vol. 74, no. 1, pp. 97-114, Feb. 2008.
- [8] E. Koutsoupias, C. H. Papadimitriou, "Worstcase equilibria", *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, pp. 404-413, 1999.
- [9] B. Awerbuch, Y. Azar, and L. Epstein, "The price of routing Unsplittable flow", *Proc.37th Symp. Theory of Computing*, pp. 57-66, 2005.
- [10] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules", *Advances in Applied Mathematics*, vol. 6, no. 1, pp. 422, 1985.
- [11] R. Agrawal, "Sample mean based index policies with $O(\log(n))$ regret for the multi-armed bandit problem", *Adv. Appl. Probab.*, vol. 27, no. 4, pp. 1054-1078, Dec. 1995.
- [12] P. Auer, N. Cesa-Bianchi, and P. Fisher, "Finite time Analysis of the Multiarmed Bandit Problem", *Machine Learning*, vol. 47, no.2-3, pp.235-256, May, 2002.
- [13] K. Liu and Q. Zhao, "Adaptive Shortest-Path Routing under Unknown and Stochastically Varying Link States", *Proc. of the 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pp. 232-237, May, 2012.
- [14] Y. Gai, B. Krishnamachari, and R. Jain, "Combinatorial Network Optimization with Unknown Variables: Multi-Armed Bandits with Linear Rewards and Individual Observations", *IEEE/ACM Transactions on Networking*, vol. 20, no. 5, 2012.
- [15] K. Liu and Q. Zhao, "Distributed Learning in Multi-Armed Bandit With Multiple Players", *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, vol. 58, no. 11, pp. 5667-5681, Nov. 2010.
- [16] Dimitri P. Bertsekas and Robert G. Gallager, "Data Networks(2nd edition)", *Prentice Hall*, 1992.
- [17] K. Liu and Q. Zhao, "Multi-Armed Bandit Problems with Heavy-Tailed Reward Distributions", *Proc. of Allerton Conference on Communications, Control, and Computing*, pp. 485-492, Sep. 2011.